

# Hotel Review Sentiment Analysis

Baraka Kautsar Sofiuddin

2023-07-11

## Importing Libraries

```
library(pacman)
p_load(pacman, qdap, ggplot2, forcats, tidyverse, dplyr, tidyr, ggthemes, wordcloud, stringr, tidytext, psych)
```

## Importing Dataset

The dataset used in this project can be found at <https://www.kaggle.com/datasets/andrewmvd/trip-advisor-hotel-reviews>  
(<https://www.kaggle.com/datasets/andrewmvd/trip-advisor-hotel-reviews>)

## Reading dataset

```
hotelReview <- read.csv("tripadvisor_hotel_reviews.csv")
head(hotelReview)
```

##

## Review

## 1

nice hotel expensive parking got good deal stay hotel anniversary, arrived late evening took advice previous reviews did valet parking, check quick easy, little disappointed non-existent view room clean nice size, bed comfortable woke stiff neck high pillows, not soundproof like heard music room night morning loud bangs doors opening closing hear people talking hallway, maybe just noisy neighbors, aveda bath products nice, did not goldfish stay nice touch taken advantage staying longer, location great walking distance shopping, overall nice experience having pay 40 parking night,

## 2 ok nothing special charge diamond member hilton decided chain shot 20th anniversary seattle, start booked suite paid extra website description not, suite bedroom bathroom standard hotel room, took printed reservation desk showed said things like tv couch ect desk clerk told oh mixed suites description kimpton website sorry free breakfast, got kidding, embassy suits sitting room bathroom bedroom unlike kimpton calls suite, 5 day stay offer correct false advertising, send kimpton preferred guest website email asking failure provide suite advertised website reservation description furnished hard copy reservation printout website desk manager duty did not reply solution, send email trip guest survey did not follow email mail, guess tell concerned guest.the staff ranged indifferent not helpful, asked desk good breakfast spots neighborhood hood told no hotels, gee best breakfast spots seattle 1/2 block away convenient hotel does not know exist, arrived late night 11 pm inside run bellman busy chatting cell phone help bags.prior arrival emailed hotel inform 20th anniversary half really picky wanted make sure good, got nice email saying like deliver bottle champagne chocolate covered strawberries room arrival celebrate, told needed foam pillows, arrival no champagne strawberries no foam pillows great room view alley high rise building good not better housekeeping staff cleaner room property, impressed left morning shopping room got short trips 2 hours, beds comfortable.not good ac-heat control 4 x 4 inch screen bring green shine directly eyes night, light sensitive tape controls.this not 4 start hotel clean business hotel super high rates, better chain hotels seattle,

## 3

nice rooms not 4\* experience hotel monaco seattle good hotel n't 4\* level.positives large bathroom mediterranean suite comfortable bed pillowsattentive housekeeping staffnegatives ac unit malfunctioned stay desk disorganized, missed 3 separate wakeup calls, concierge busy hard touch, did n't provide guidance special requests.tv hard use ipod sound dock suite non functioning. decided book mediterranean suite 3 night weekend stay 1st choice rest party filled, comparison w spent 45 night larger square footage room great soaking tub whirlpool jets nice shower.before stay hotel arrange car service price 53 tip reasonable driver waiting arrival.checkin easy downside room picked 2 persons jacuzzi tub no bath accessories salts bubble bath did n't stay, night got 12/1a checked voucher bottle champagne nice gesture fish waiting room, impression room huge open space felt room big, tv far away bed chore change channel, ipod dock broken disappointing.in morning way asked desk check thermostat said 65f 74 2 degrees warm try cover face night bright blue light kept, got room night no, 1st drop desk, called maintenance came look thermostat told play settings happy digital box w o n't work, asked wakeup 10am morning did n't happen, called later 6pm nap wakeup forgot, 10am wakeup morning yep forgotten.the bathroom facilities great room surprised room sold whirlpool bath tub n't bath amenities, great relax water jets going,

## 4

unique, great stay, wonderful time hotel monaco, location excellent short stroll main downtown shopping area, pet friendly room showed no signs animal hair smells, monaco suite sleeping area big striped curtains pulled closed nice touch felt cosy, goldfish named brandi enjoyed, did n't partake free wine coffee/tea service lobby thought great feature, great staff friendly, free wireless internet hotel worked suite 2 laptops, decor lovely eclectic mix patterns color palette, animal print bathrobes feel like rock stars, nice did n't look like sterile chain hotel hotel personality excellent stay,

## 5

great stay great stay, went seahawk game awesome, downfall view building did n't complain, room huge staff helpful, booked hotels website seahawk package, no charge parking got voucher taxi, problem taxi driver did n't want accept voucher barely spoke english, funny thing speak arabic called s

```

tartet making comments girlfriend cell phone buddy, took second realize just said fact speak langu
age face priceless, ass told, said large city, told head doorman issue called cab company promptly
answer did n't, apologized offered pay taxi, bucks 2 miles stadium, game plan taxi return going hu
mpin, great walk did n't mind, right christmas wonderful lights, homeless stowed away building ent
rances leave, police presence not greatest area stadium, activities 7 blocks pike street waterfron
t great coffee shops way, hotel mantained foyer awesome, wine tasting available evening, best dog,
taking st. bernard time family, safes hotel located service desk room, bathroom huge jetted tub hu
ge, funny house keeping walked girlfriend getting dressed, did n't hear knock doing turn service,
screamed girlfriend screams hit floor laughing, started talking spanish worked, place recommend pr
ice, check online deals just good not better, besite contains deals vouchers travel websites n't t
ell,
## 6
love monaco staff husband stayed hotel crazy weekend attending memorial service best friend husban
d celebrating 12th wedding anniversary, talk mixed emotions, booked suite hotel monte carlos, loan
ed beautiful fan-tanned goldfish named joliet weekend visited dogs worked desk human companions, r
oom decorated nicely couch used pillows, l'occitane bath amenities welcome sight, room quiet peace
ful, wireless internet access wonderful server went morning leaving problems printing boarding pas
ses, afternoon reception serves oenophile-satisfying wine australia scrumptious cookies, restauran
t closed renovation stay finally ate food good drinks better, word caution restaurant larger perso
n not sit booths wo n't fit, 5'6 125 lbs husband 5'9 175. table smack-against stomach couple inche
s space mighty uncomfortable patron larger pregnant, bad design opinion place decorated funky welc
oming way metal wood handblown glass light fixtures expect seattle capital glass art industry, def
initely stay reason,
## Rating
## 1 4
## 2 2
## 3 3
## 4 5
## 5 5
## 6 5

```

# Exploring The Data

## Overview & Review Data Deepdive

```

describe(hotelReview)

##          vars      n    mean      sd median  trimmed      mad min    max range
## Review*      1 20491 10246.00 5915.39  10246 10246.00 7595.36    1 20491 20490
## Rating       2 20491    3.95    1.23     4    4.15    1.48    1     5     4
##              skew kurtosis      se
## Review*    0.00      -1.2 41.32
## Rating   -1.08      0.1  0.01

#characters
hotelReview%>%
  summarize(mean_character = mean(nchar(Review)), median_character = median(nchar(Review)))

##      mean_character median_character
## 1           724.8968             537

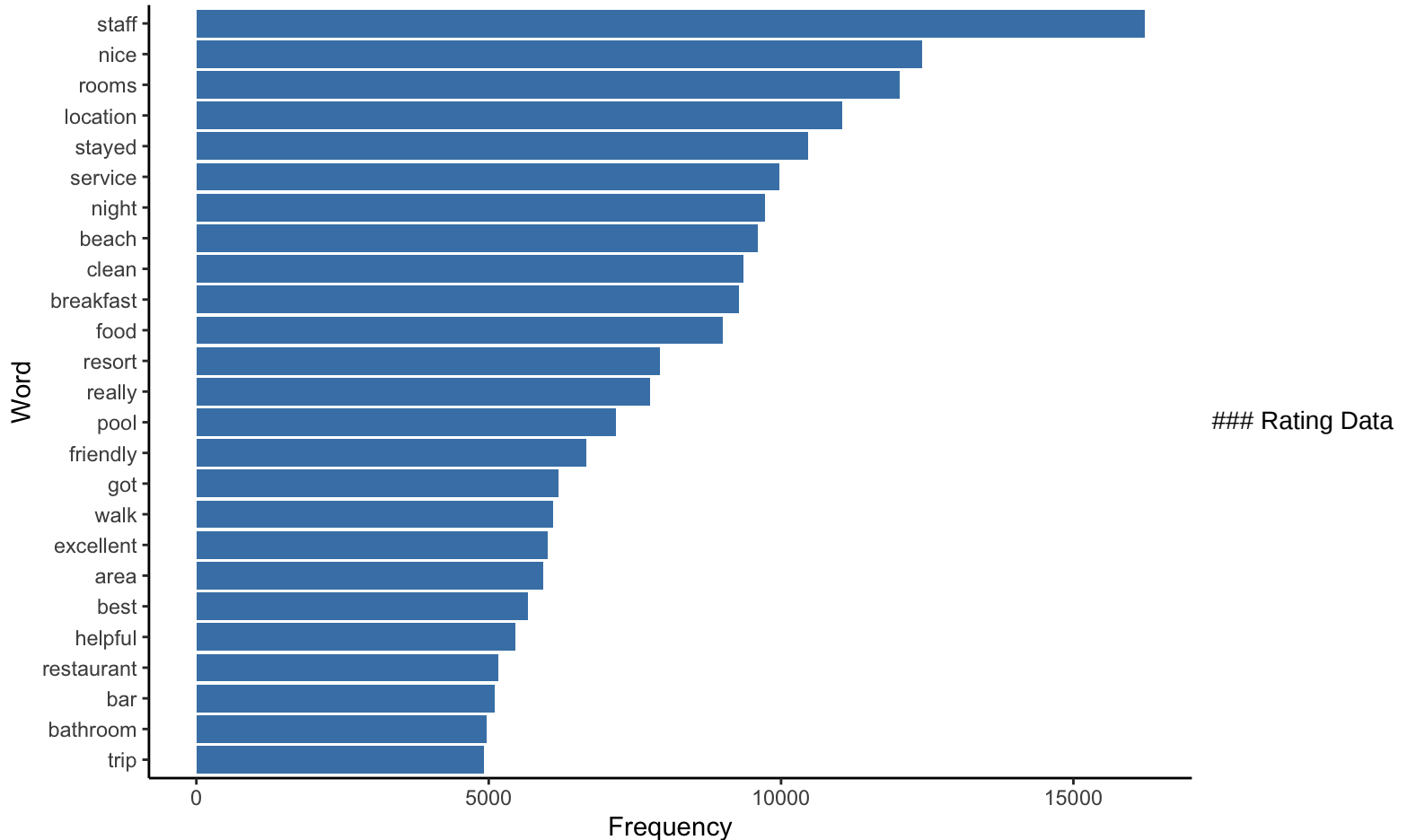
```

```
#words
hotelReview%>%
  summarize(mean_words = mean(str_count(string = Review,pattern = '\\S+')), median_words = median
(str_count(string = Review,pattern = '\\S+')))
```

```
##      mean_words median_words
## 1      104.3758           77
```

```
most_common_word = freq_terms(text.var=hotelReview$Review,top=25,stopwords =
                             c(Top200Words,tm::stopwords("english"),"hotel","room","sta
y","n't"))
```

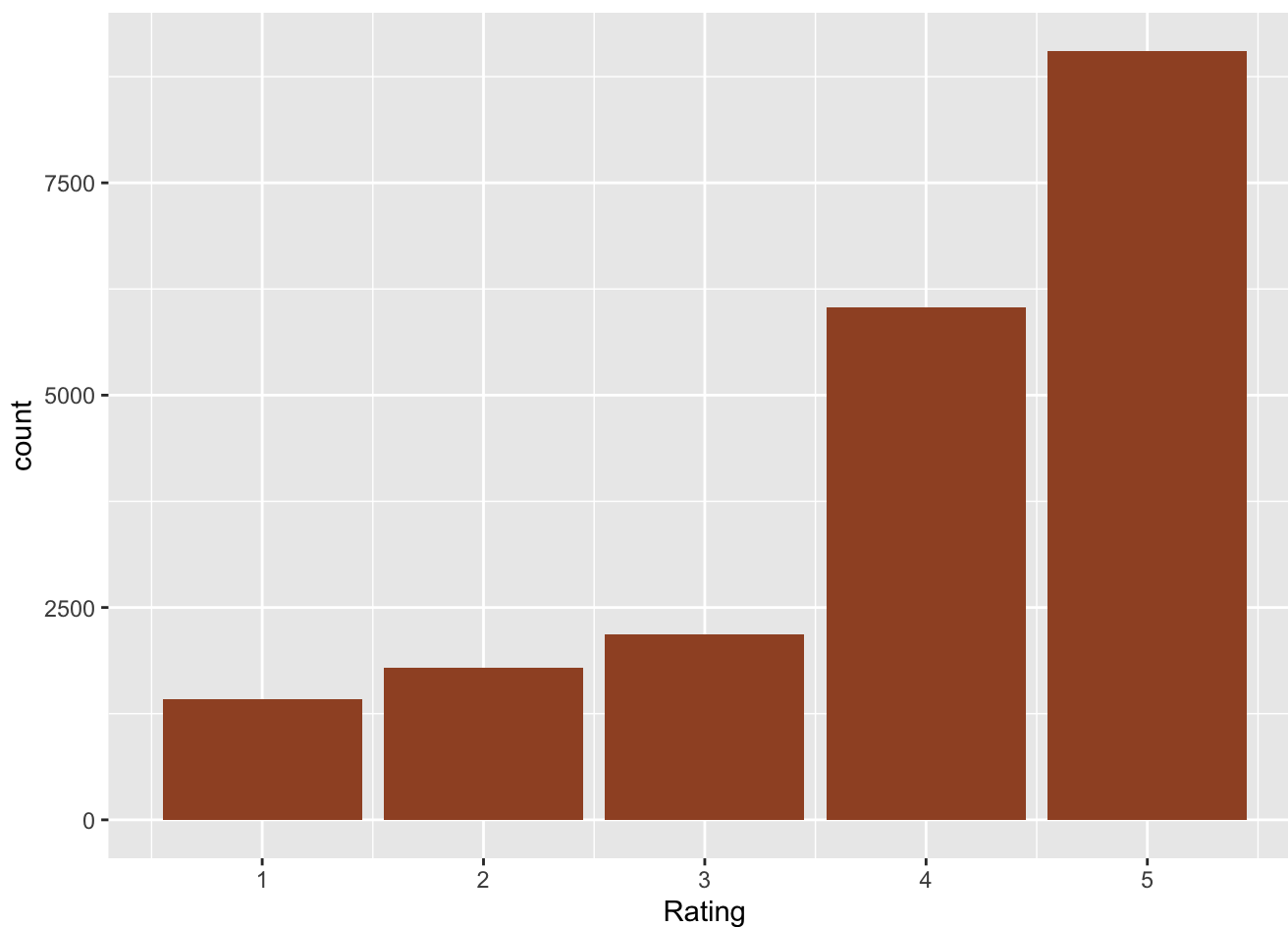
```
ggplot(most_common_word,aes(x = fct_reorder(WORD, FREQ, .desc=FALSE),y=FREQ))+
  geom_bar(stat = "identity",fill = "steelblue")+
  xlab("Word")+
  ylab("Frequency")+
  theme_classic()+
  coord_flip()
```



```
table(hotelReview$Rating)
```

```
##
##      1      2      3      4      5
## 1421 1793 2184 6039 9054
```

```
ggplot(data=hotelReview,aes(x=Rating))+  
  geom_bar(fill='sienna')
```



## Sentiment Analysis

### Word Lexicon

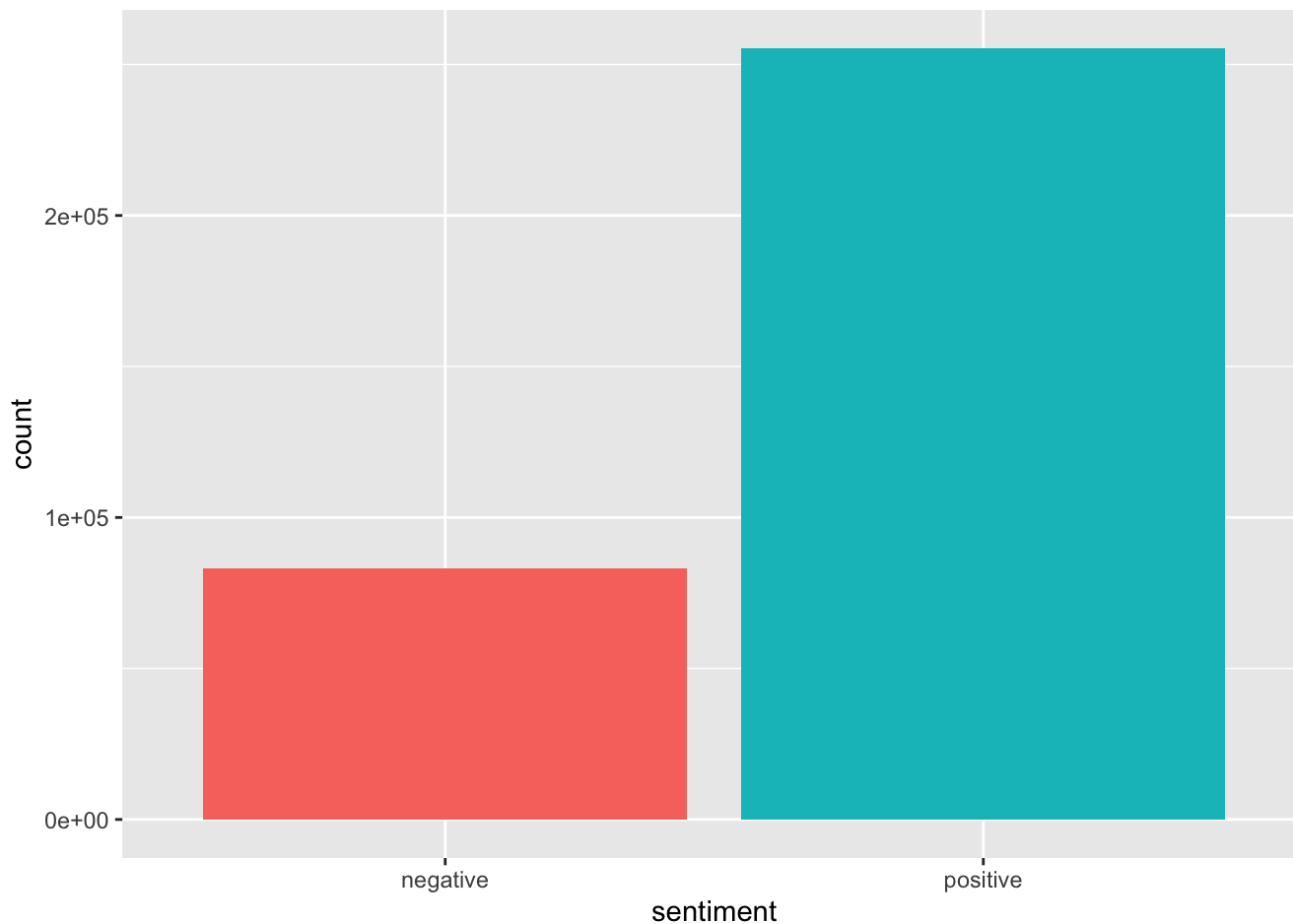
```
pos_neg = hotelReview%>%  
  select(Review)%>%  
  unnest_tokens(output = word, input = Review)%>%  
  inner_join(get_sentiments('bing'))
```

```
## Joining with `by = join_by(word)`
```

```
head(pos_neg)
```

```
##      word sentiment  
## 1    nice  positive  
## 2 expensive negative  
## 3    good  positive  
## 4    easy  positive  
## 5 disappointed negative  
## 6    clean  positive
```

```
ggplot(pos_neg, aes(x=sentiment, fill=sentiment))+
  geom_bar()+
  guides(fill = F)
```



```
options(scipen=999)
```

## sentiment proportion for each rating category

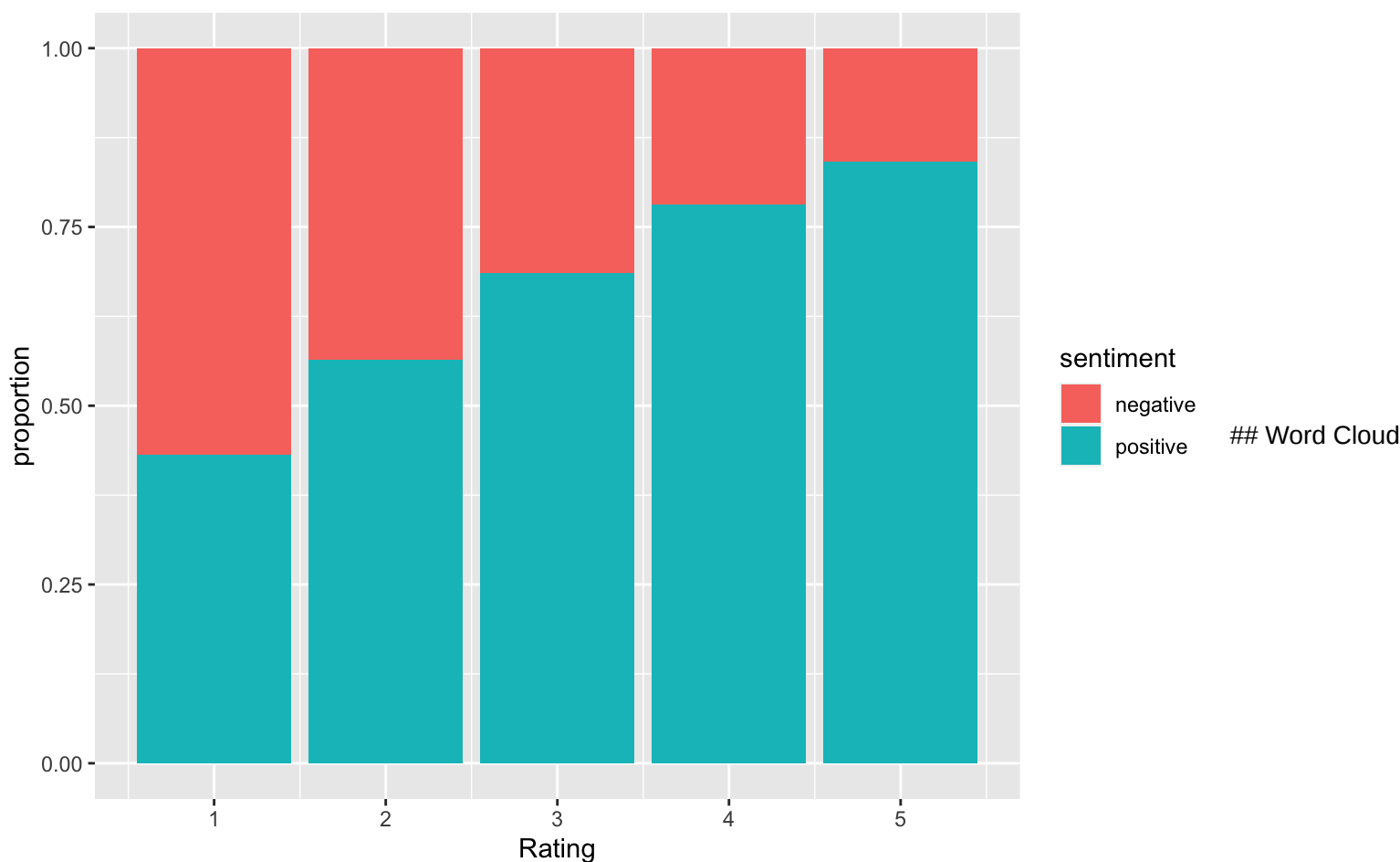
```
rating_sentiment =
  hotelReview %>%
    select(Review, Rating)%>%
    unnest_tokens(output=word, input=Review)%>%
    inner_join(get_sentiments('bing'))%>%
    group_by(Rating, sentiment)%>%
    summarize(amount = n())%>%
    mutate(proportion = amount/sum(amount))
```

```
## Joining with `by = join_by(word)`
```

```
## Warning in inner_join(., get_sentiments("bing")): Detected an unexpected many-to-many relationship
## between `x` and `y`.
## i Row 1032757 of `x` matches multiple rows in `y`.
## i Row 4229 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
## "many-to-many"` to silence this warning.
```

```
## `summarise()` has grouped output by 'Rating'. You can override using the
## `.groups` argument.
```

```
rating_sentiment %>%
  ggplot(aes(x=Rating,y=proportion,fill=sentiment))+
  geom_col()
```



```
wordcloudData =
  hotelReview%>%
  select(Review)%>%
  unnest_tokens(output=word,input=Review)%>%
  anti_join(stop_words)%>%
  inner_join(get_sentiments('bing'))%>%
  count(sentiment,word,sort=T)%>%
  spread(key=sentiment,value = n,fill=0)%>%
  data.frame()
```

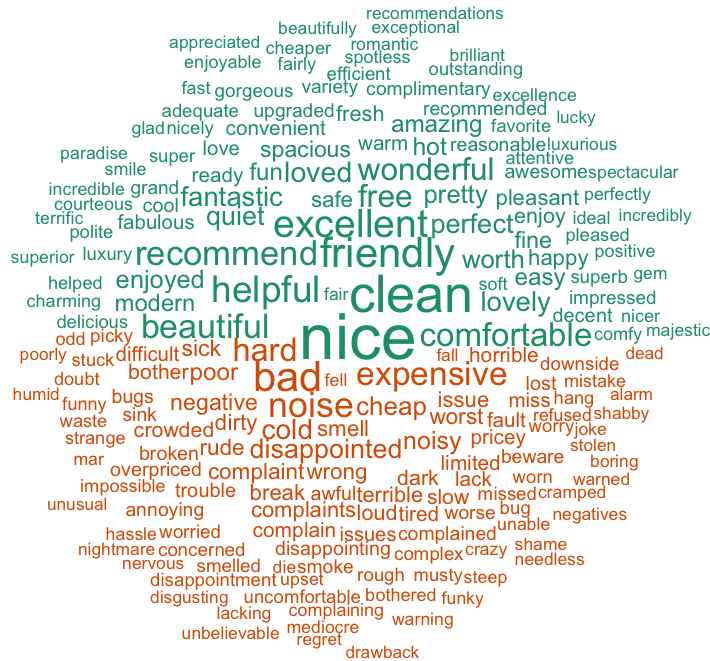
```
## Joining with `by = join_by(word)`
## Joining with `by = join_by(word)`
```

```
## Warning in inner_join(., get_sentiments("bing")): Detected an unexpected many-to-many relationship
## between `x` and `y`.
## i Row 810002 of `x` matches multiple rows in `y`.
## i Row 4229 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
## "many-to-many"` to silence this warning.
```

```
rownames(wordcloudData) = wordcloudData[, 'word']
wordcloudData = wordcloudData[, c('positive', 'negative')]

set.seed(42)
comparison.cloud(term.matrix = wordcloudData, scale = c(2, 0.5), max.words = 200, rot.per=0)
```

positive



negative

```
p_unload(all)
```

```
## The following packages have been unloaded:
```

```
## psych, tidytext, wordcloud, ggthemes, lubridate, stringr, dplyr, purrr, readr, tidyr, tibble,  
tidyverse, forcats, ggplot2, qdap, RColorBrewer, qdapTools, qdapRegex, qdapDictionaries, pacman
```